

Effect of Partially Correlated Data on Clustering in Wireless Sensor Networks

Mahdi Lotfinezhad and Ben Liang
Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, Ontario, M5S 3G4, Canada
Email: {mlotfinezhad, liang}@comm.utoronto.ca

Abstract—In wireless sensor networks, clustering allows the aggregation of sensor data. It is well known that leveraging the correlation between different samples of the observed data will lead to better utilization of energy reserve. However, no previous work has analyzed the effect of non-ideal data aggregation in multi-hop sensor networks. In this paper, we propose a novel analytical framework to study how partially correlated data affect the performance of clustering algorithms. We analyze the behavior of multi-hop routing and, by combining random geometry techniques and rate distortion theory, predict the total energy consumption and network lifetime. We show that when a moderate amount of correlation is available, the optimal probabilities that lead to minimum energy consumption are far from optimality in terms of network lifetime. In addition, we study the sensitivity of the total energy consumption and network lifetime to the amount of correlation and compression distortion constraint.

I. INTRODUCTION

Recent advances in wireless communications and rapid development in small low-cost microelectronic and micro-electromechanical sensor technology along with the advances in wireless sensors have made it possible to deploy large number of sensors to form a wireless sensor network. Wireless sensor networks can be used for various purposes, e.g., vehicle tracking and habit monitoring [1][2][3]. For large scale sensor networks to be practical, the cost and size of these sensors are of primary concern. As a result, they are usually equipped with small batteries. This constraint has significant effect on the network design and makes energy efficiency a major design challenge.

Clustering improves the scalability of multi-hop wireless networks. Typically, a clustering algorithm divides the network into subsets of nodes, called clusters, each with one node serving as its cluster-head (CH). After the formation of clusters, nodes transmit their data to the CHs for data aggregation. In their pioneering work, the authors of [4] have studied the capacity of wireless networks and proved that the capacity per node diminishes as the network size increases. However, recently it has been shown that, due to data correlation, large-scale multi-hop sensor networks are feasible [5]. Clearly, clustering provides an architectural framework for exploiting data correlation.

Various clustering protocols have been proposed, either in the context of generic wireless ad hoc networks [6][7][8][9][10][11] or wireless sensor networks [12][13][14][15]. These protocols either do not consider data correlation or assume ideal data aggregation, where data are perfectly correlated, such that an arbitrary number of packets within a cluster can be compressed to one packet. However, in practical sensor networks, the performance of data aggregation is closely related to the various levels of data correlation. This necessitates additional study into the characteristics of clustering with partially correlated data.

In this paper, we consider the effect of partially correlated data on the performance of clustering algorithms. As far as we are aware, this is the first paper that provides a comprehensive analytical framework to evaluate the energy and lifetime performance of clustering in sensor networks. The proposed analysis is generic and can be applied to a wide array of random clustering algorithms.

In particular, we apply random geometry methodologies [16] to analyze the energy consumption for forwarding data in a multi-hop sensor network. Combining this result with rate distortion theory [17], we provide a mathematical analysis framework to study the energy consumption and lifetime of a network when arbitrary amount of correlation is present. We study the imbalance in the energy consumption of sensors as a function of their position, through which we then show the trade-off between total energy consumption and network lifetime and conclude that optimal network energy consumption necessarily does not lead to an optimal network lifetime. In addition, we present numerical analysis results that illustrate the relation between data correlation and compression distortion, and energy consumption and network lifetime.

The rest of this paper is organized as follows. In Section II, we define the system model used in our analysis and simulation. In Sections III, IV, V, and VI, we present our mathematical analysis on data forwarding, sensor data correlation, network energy consumption, and network lifetime, respectively. We justify our analysis via simulation and study the effect of data correlation on energy consumption and network lifetime in Section VII. In Section VIII, we describe the related work in detail. Finally, we conclude the paper in Section IX.

II. SYSTEM MODEL

To clarify our assumptions and network setup, we present the system model in two subsections as follows.

A. Node Deployment and Characteristics

Our network consists of a set of sensors and one processing center or data sink. A total of N sensors are randomly distributed with a density of λ in an area A . If the area A and the total number of sensors is large enough, then this point process can be approximated by a homogeneous Poisson point process with density λ . We assume that the processing center is located at the center of the network and has unlimited energy supply. We also assume that all sensors in the network as well as the processing center are stationary. Our study also applies to sensor networks with multiple data sinks, each in charge of a fixed set of sensors. In this case, the performance of the clustering can be analyzed independently for each subset of the network surrounding a data sink.

Sensors are aimed to be extremely small, inexpensive, and simple devices. Therefore, all sensors within the network are assumed to transmit at a fixed transmission power, and each sensor has the same radio range R . All sensors are equipped with a battery that has an initial amount of energy equal to E_0 . We assume that each sensor requires 0.5 units of energy to transmit or receive one unit of data. As will be shown later, the exact ratio of transmitting to receiving power will not affect our analysis. We also assume that the communication environment is contention- and error-free, and therefore sensors do not have to retransmit any data. The choice and effect of centralized scheduling or distributed MAC are outside the scope of this work.

B. Underlying Routing and Clustering Protocols

Throughout this paper, we use the terms ‘‘sensor’’ and ‘‘node’’ interchangeably. To forward data to CHs, and from CHs to the processing center, we assume that nodes use light weight Minimum Hop Routing (MHR). The advantage of MHR is two fold. First, it matches well with the fixed transmission power of inexpensive sensors. Second, since the sensors are stationary, MHR requires very infrequent route updating and hence much less energy consumption than other more active routing protocols such as energy-based routing. In our approach, nodes may not be aware of the topology of the network, and the paths can be setup when sensor information requests are disseminated by the data sink, similar to the approach taken in [18].

We assume a generic clustering algorithm where CHs are selected randomly. One example of implementing a random clustering algorithm is presented in [14] in which CHs are selected uniformly throughout the network. The clustering algorithm is run every T units of time, where T generally depends on the type of application and initial energy supply. Each T units of time is divided into M rounds, in each of which, a CH schedules nodes within its cluster and receives observed data from them. In the beginning of each T rounds,

every node selects itself as a CH with a fixed probability of p . It can be easily shown that the CHs form a Poisson point process with density $\lambda_1 = p\lambda$, and the subset of nodes that are not CHs will form another Poisson point process with density $\lambda_0 = (1 - p)\lambda$. After the CHs have been selected, each CH sends a beacon that is flooded up to k_h hops to advertise its status as a CH. Every node that is not a CH will join the cluster whose CH is the nearest (by using hop count information as the first approximation to distance). This process tessellates the network to a set of Voronoi cells.

As derived in [14], one can put an upper limit on k_h which can be derived from the following formula.

$$k_h = \lceil \frac{1}{R} \sqrt{\frac{-0.917 \ln(\alpha/7)}{p\lambda}} \rceil.$$

With the specified value of k_h , the probability that the radius of the minimum ball centered at the nucleus of clusters (CHs) is bigger than k_h hops, is less than α . We typically set $\alpha = 0.001$ to ensure that most nodes can receive the beacon packet from their corresponding nearest CH within k_h hops. However, it is possible that a node doesn't receive a beacon from nearby CHs. In this case, it will select itself as a forced CH [14]. It is easy to see that cluster formation process has a time complexity of $O(k_h)$. After cluster formation, all CHs will schedule their sensors to begin communication. The exact scheduling algorithm is out of the scope of this paper and is a topic for future research.

We emphasize the fact that the proposed analytical framework does not depend on the clustering protocol details. Other possible implementation include those listed in section VIII.

In our model, each node observes a sample of the measured random field and sends its own sample to the assigned CH. After the reception of all data samples from its cluster-members, the CH performs data aggregation subject to a total distortion of D and sends the aggregated data to the data sink through multi-hop routing over all sensor nodes. We consider the worst case where nodes send each of their data samples with the same amount of bits, giving a distortion of D_0 , as if there were no aggregation, and consequently aggregation is done only in CHs. This amount of distortion, D_0 , equals the average distortion per sample, which is obtained after data compression in CHs.

In the sequel we present analysis on the performance of sensor clustering in terms of energy consumption and network lifetime.

III. ENERGY CONSUMPTION ON DATA FORWARDING

For our analysis we use the following definitions.

- The sensing field is assumed to be a circular disc with radius of KR , for some integer K .
- Nodes are distributed with density λ .
- We define $E_c(r)$ as the conditional expected value of energy consumption of a node as a function of its distance to the processing center, i.e., if a given node is at radial distance r from the processing center, then $E_c(r)$ is the expected value of energy consumption of that node.

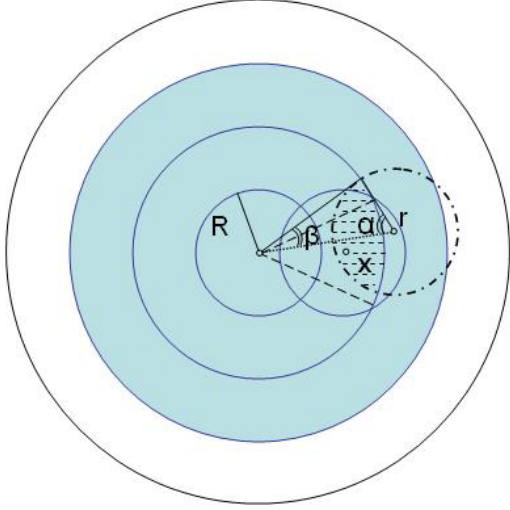


Fig. 1. Sensor network divided into layers

- Similarly, we define $N_c(r)$ as the expected number of packets that a node at radial distance r from the processing center must forward including its own packet.

Fig. 1 shows the sample topology that will be used for our analysis. If the density of nodes is sufficiently high such that there are enough paths to select from, then we expect the number of hops needed to reach the processing center to be approximately $\lceil \frac{r}{R} \rceil$ when MHR is used. This simple observation allows us to consider the sensing field as consisting of different layers, each with a thickness of R as shown in Fig. 1.

From here on, we use the terms point and node interchangeably. Associated with any given point \vec{x} we draw a circle showing its transmission range. In the example of Fig. 1, \vec{x} is in the second layer. Only those points which are in the intersection of the transmission range of \vec{x} and the third layer can be potential nodes that will select \vec{x} as their next hop. Denote one such point \vec{r} . To find the probability that this occurs (i.e., when \vec{r} selects \vec{x} as the next hop), select point \vec{y} as a point which is in the transmission range of \vec{r} and is in the second layer. If points \vec{x} and \vec{y} are in the same layer, then from the point of view of point \vec{r} , there is not considerable distinction between points \vec{x} and \vec{y} . As a special case, if only points \vec{x} and \vec{y} are present in the transmission range of \vec{r} , then point \vec{r} will choose one of these points with equal probability.

To generalize this idea, let \vec{r} be a point in the k_{th} layer and within the transmission range of \vec{x} , and let $A(\vec{x}, \vec{r})$ be the intersection of circle centered at \vec{r} with a radius of R and the circle centered at the processing center with radius $(k-1)R$. Therefore $A(\vec{x}, \vec{r})$, which is denoted by the dashed-shaded region in Fig. 1, shows the potential region for the next hop of point \vec{r} . Following the reasoning stated earlier, the probability that the node at point \vec{r} sends its packet to the

point \vec{x} (i.e., chooses \vec{x} as the next hop) is

$$\begin{aligned} p(\vec{x}, \vec{r}) &= \sum_{n=1}^{\infty} \frac{1}{n} Pr(N_A(\vec{x}, \vec{r}) = n | N_A(\vec{x}, \vec{r}) \geq 1) \\ &= \sum_{n=1}^{\infty} \frac{e^{-\lambda A(\vec{x}, \vec{r})}}{1 - e^{-\lambda A(\vec{x}, \vec{r})}} \frac{(\lambda A(\vec{x}, \vec{r}))^n}{n!n}, \end{aligned} \quad (1)$$

where N_A denotes the number of nodes in $A(\vec{x}, \vec{r})$, and we have used the fact that the probability of having n nodes in an area of A is $\frac{(\lambda A)^n}{n!} e^{-\lambda A}$. To find $A(\vec{x}, \vec{r})$, we make use of the cosine law to find the angles β and α

$$\begin{aligned} \beta &= \arccos\left(\frac{(\lceil \frac{|\vec{x}|}{R} \rceil R)^2 + |\vec{r}|^2 - R^2}{2|\vec{r}|(\lceil \frac{|\vec{x}|}{R} \rceil R)}\right) \\ \alpha &= \arccos\left(\frac{R^2 + |\vec{r}|^2 - (\lceil \frac{|\vec{x}|}{R} \rceil R)^2}{2R|\vec{r}|}\right), \end{aligned}$$

where $|\vec{x}|$ denotes distance from \vec{x} to the data sink, and thus $A(\vec{x}, \vec{r})$ is given by

$$\begin{aligned} A(\vec{x}, \vec{r}) &= \beta \left(\lceil \frac{|\vec{x}|}{R} \rceil R\right)^2 - \left(\lceil \frac{|\vec{x}|}{R} \rceil R\right)^2 \frac{\sin(2\beta)}{2} \\ &\quad + \alpha R^2 - R^2 \frac{\sin(2\alpha)}{2}. \end{aligned} \quad (2)$$

To determine the expected total number of packets that a node at point \vec{x} would send, we need to calculate a recursive integration. This integration is over the intersection of the transmission range of \vec{x} and the k_{th} layer, which leads to the following.

$$\begin{aligned} N_c(|\vec{x}|) &= 1 + \int_{\lceil \frac{|\vec{x}|}{R} \rceil R}^{R+|\vec{x}|} \int_{-\gamma}^{\gamma} \frac{e^{-\lambda A(\vec{x}, r)}}{1 - e^{-\lambda A(\vec{x}, r)}} \\ &\quad \left(\sum_{n=1}^{\infty} \frac{(\lambda A(\vec{x}, r))^n}{n!n} \right) N_c(r) \lambda d\theta r dr, \end{aligned} \quad (3)$$

where

$$\gamma = \arccos\left(\frac{|\vec{x}|^2 + r^2 - R^2}{2r|\vec{x}|}\right).$$

In the above, $N_c(|\vec{x}|)$ accounts for all of the paths that go through \vec{x} . We add one to the integration because we have assumed that every node has its own packet to send. The relationship between $N_c(\vec{x})$ and $E_c(\vec{x})$ is

$$E_c(\vec{x}) = N_c(\vec{x}) - 0.5,$$

where we have assumed that both transmission and reception of a packet require .5 units of energy for a total of 1 unit. Every node which relays a packet consumes exactly one unit of energy except for when a node is transmitting its own packet. Clearly, our analysis is relatively independent of the ratio of transmitting and receiving energy consumption.

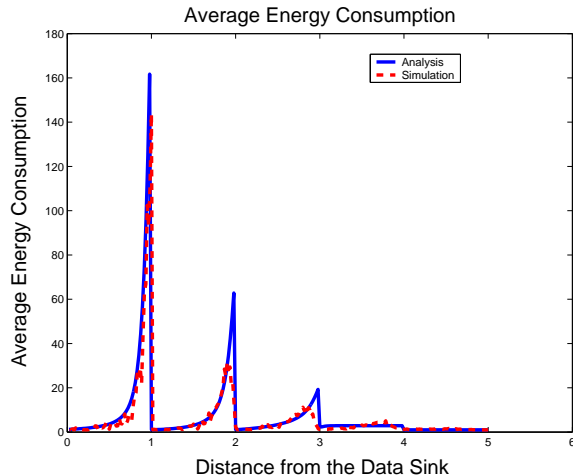


Fig. 2. Energy consumption of nodes as a function of distance

To confirm the above analysis results, we have simulated a network with a diameter of $5R$ and node density $\frac{100}{\pi R^2}$. We run the simulation for 200 different networks with randomly distributed nodes and take their average. Fig. 2 shows that the analysis matches the results obtained via simulation.

From Fig. 2, we observe that there are very sharp peaks in the energy consumption curve, which emphasizes the fact that nodes at particular distances will die first even though they are farther away from the processing center than other nodes in the network. This discontinuity is a unique feature of minimum hop routing and it may seem counterintuitive initially. In fact, nodes that are close to the outer edge of each layer are seen by more nodes in the next layer. Consequently, these nodes have a greater chance to be selected as relays compared with the nodes that are within the same layer but are closer to the inner edge of that layer.

For this reason, it can be advantageous to have heterogenous nodes in terms of initial energy supply, whose E_0 is a function of distance to the data sink such that it is proportional to the expected energy consumption as in Fig. 2. That will result in a flat profile for the residual energy of all nodes. This approach can be useful especially when the network has a small diameter. The high peak to average ratio of the energy consumption curve means that we need to place few nodes with high E_0 at the edge of each layer. Alternatively, nodes with equal E_0 can be deployed non-uniformly according to the obtained energy consumption curve to achieve a flat energy profile.

IV. SENSOR DATA CORRELATION

Let $\mathbf{X} = \{x_i, i = 1, 2, \dots, n\}$ be a vector of n samples of the measured random field returned by the n sensors within a cluster. Let $\hat{\mathbf{X}}$ be a representation of \mathbf{X} , and $d(\hat{\mathbf{X}}, \mathbf{X})$ be a distortion measure. It has been shown that the minimum number of bits required to represent \mathbf{X} subject to a bound on the total distortion, i.e., $E(d(\hat{\mathbf{X}}, \mathbf{X})) \leq D$, can be computed by the

following formula [19]

$$R(D) = \min_{f(\hat{\mathbf{X}}|\mathbf{X}): E(d(\mathbf{X}, \hat{\mathbf{X}})) \leq D} I(\mathbf{X}, \hat{\mathbf{X}}), \quad (4)$$

where $I(\mathbf{X}, \hat{\mathbf{X}})$ is the mutual information between \mathbf{X} and $\hat{\mathbf{X}}$. When the distortion measure is mean square error (MSE), i.e., $d(\mathbf{X}, \hat{\mathbf{X}}) = \|\mathbf{X} - \hat{\mathbf{X}}\|^2$, a Gaussian source is the worst case and needs most bits to be represented compared with the other sources [5]. Therefore for the purpose of illustration, we consider the case where \mathbf{X} is a multivariate Gaussian random vector, i.e., $\mathbf{X} \sim \mathcal{N}(\bar{\mathbf{M}}, \Sigma)$. We assume that the correlation between two samples is a decreasing function of the Euclidian distance between the sensors where the samples are taken. It has been shown that, for MSE as the distortion measure and a constraint on the total amount of distortion

$$E(\|\mathbf{X} - \hat{\mathbf{X}}\|^2) \leq D,$$

the minimum number of bits required to represent data for a Gaussian source, \mathbf{X} in our case, is [17]

$$R(D) = \sum_{n=1}^N \frac{1}{2} \log \frac{\lambda_n}{D_n}, \quad (5)$$

where $\lambda_1 \geq \lambda_2 \dots \geq \lambda_N$ are the ordered eigenvalues of the correlation matrix Σ and

$$\sum_{n=1}^N D_n = D, \\ D_n = \begin{cases} \theta, & \text{if } \theta < \lambda_n \\ \lambda_n, & \text{otherwise,} \end{cases}$$

or equivalently θ is chosen such that

$$\sum_{n=1}^N \min(\theta, \lambda_n) = D. \quad (6)$$

From (6) it can be easily shown that for reasonable amounts of distortion D where $\sum_{n=1}^N \lambda_n \geq D$, there exists $n_0 \leq N$ such that

$$\theta = \frac{D - \sum_{n=n_0+1}^N \lambda_n}{n_0}, \\ \lambda_{n_0+1} \leq \theta < \lambda_{n_0}.$$

Also, (5) shows that the rate distortion in the case of a Gaussian random field is only a function of the eigenvalues of the correlation matrix. This result is well known as ‘‘reverse water-filling.’’

Later in our analysis, we need to know the amount of compression as a function of the number of samples available. Although it is possible to derive some asymptotic results for $R(D)$ in a random topology [5], these asymptotic results are not sufficient for our purposes, especially when the number of samples is not high enough. Therefore, we have used the Monte Carlo method whose input is the random node locations subject to node density λ . For a given number of nodes N , we consider a disk with an area of $A = (\frac{N}{\lambda})$ and distribute nodes randomly in it. We derive the number of bits according to (5) and find

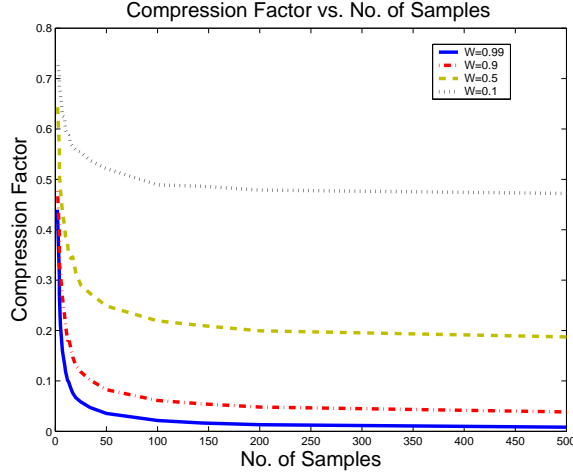


Fig. 3. Compression factor for $\lambda = 6.25 m^{-2}$

its mean value over different scenarios. The results are shown as the compression factor in Fig. 3. As an example, we have considered the case where the correlation falls as a Gaussian function which can be converted to $f(p_i, p_j) = \sigma^2 W \|p_i - p_j\|^2$, where p_i and p_j are two points in the 2D plane, and $\|p_i - p_j\|$ is their Euclidean distance. W is the representative of the amount of correlation between spatial samples and should be less than one such that Σ is a semi-positive definite matrix. Here σ^2 is a constant and denotes the variance of each sample of the measured field. We choose $\sigma^2 = 1$ for brevity.

As Fig. 3 shows, as W decreases the correlation also decreases, which means a lower amount of compression can be achieved. Also it can be seen that the compression factor tends to be a constant as the number of nodes is increased. This is an intuitive result, because as we increase the number of samples, samples that are really far apart have very low correlation, and effectively they can be considered as independent data.

V. TOTAL ENERGY CONSUMPTION

To study the total energy consumption we need to know 1) the average amount of total energy needed to communicate with the CHs and 2) the average amount of total energy that the CHs need to communicate with the data sink.

As a first approximation to the routing performance, we need to know the average value of $\lceil \frac{r}{R} \rceil$ as the average number of hops to the CHs or the average number of hops from the CHs to the processing center. Let \mathcal{C}_0 represent a typical Voronoi cell whose nucleus is located at the origin, Π_0 represent the Poisson point process associated with the non-CH nodes, and x_i be a member of Π_0 , we can define a function $f(x_i)$ as a property of x_i , e.g., its distance to the CH, and S_f as the summation of that property over all cluster-members, i.e.,

$$S_f = \sum_{x_i \in \Pi_0} f(x_i) \mathbf{1}\{x_i \in \mathcal{C}_0\}, \quad (7)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. From [16] we can compute the expected value of S_f , $\mathbf{E}S_f$, as follows

$$\mathbf{E}S_f = \lambda_0 \int f(x) \exp(-\lambda_1 \pi |x|^2) dx. \quad (8)$$

In our case $f(x)$ is a function from the 2-D plane to a real number which is $\lceil \frac{r}{R} \rceil$. Using (8) and assuming that R_{net} is the network radius we have

$$\begin{aligned} \mathbf{E}S_{\lceil \frac{r}{R} \rceil} &= \lambda_0 \int \lceil \frac{r}{R} \rceil \exp(-\lambda_1 \pi |r|^2) r dr d\theta \\ &\simeq \frac{\lambda_0}{\lambda_1} \left[\sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2) - \left(\lfloor \frac{R_{net}}{R} \rfloor + 1 \right) \exp(-\lambda_1 \pi (R_{net}^2)) \right] \\ &\simeq \frac{\lambda_0}{\lambda_1} \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2). \end{aligned} \quad (9)$$

We denote by C_1 the total energy needed to communicate with the CHs. We compute the expected value of C_1 by conditioning it by the number of clusters in the network which leads to

$$\begin{aligned} E(C_1) &= E[E(C_1 | n = n_0)] = E(n_0 \mathbf{E}S_{\lceil \frac{r}{R} \rceil}) \\ &\simeq Np \frac{\lambda_0}{\lambda_1} \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2). \end{aligned} \quad (10)$$

We denote by C_2 the amount of total energy consumed by the CHs to communicate with the data sink. Let n_c be a random variable representing the number of CHs, r_i be the distance from the i_{th} CH to the data sink, c_i be the i_{th} cluster, and $R_D(c_i)$ be the number of bits required to represent all the data samples within the i_{th} cluster, normalized to the number of bits that a sensor needs to represent its data with a distortion of D_0 , i.e., $R_D(c_i) = \frac{R(D_i)}{0.5 \log \frac{1}{D_0}}$, we have

$$\begin{aligned} E(C_2) &= E\left(\sum_{i=1}^{n_c} \lceil \frac{r_i}{R} \rceil R_D(c_i)\right) \\ &= E[n_c E(\lceil \frac{r_i}{R} \rceil) E(R_D(c_i))] \\ &= E(n_c) E(\lceil \frac{r_i}{R} \rceil) E(R_D(c_i)) \\ &= Np E(\lceil \frac{r_i}{R} \rceil) E(R_D(c_i)), \end{aligned} \quad (11)$$

where

$$E(\lceil \frac{r_i}{R} \rceil) = \int_A \frac{1}{A} \lceil \frac{\sqrt{x^2 + y^2}}{R} \rceil dA.$$

For each cluster we assume a total distortion of $D_i = N_i D_0$, where N_i represents the number of nodes in the cluster c_i . Therefore, $0.5 \log \frac{1}{D_0}$ is one unit of data in our setup. The $E(\lceil \frac{r_i}{R} \rceil)$ term depends on the geometry of the network and can be numerically computed. $E(R_D(c_i))$ depends on the specific type of correlation function and the joint pdf of random samples whose characteristics have been shown earlier.

Therefore the average total energy consumption is

$$\begin{aligned}
E(C_t) &= E(C_1) + E(C_2) \\
&\simeq Np \frac{\lambda_0}{\lambda_1} \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2) \\
&+ Np E(\lceil \frac{r_i}{R} \rceil) E(R_D(c_i)). \quad (12)
\end{aligned}$$

VI. NETWORK LIFETIME

Network lifetime, which is tightly coupled with energy consumption of nodes in the network, can be defined as the time elapsed until the first (or the last) node depletes its energy reserve [15]. To study the behavior of network lifetime as the time that the first node dies we use the results obtained in Section III. Although there are other definitions of lifetime, we believe that for the given network model our definition is a reasonable indicator of network lifetime. Given that traffic is uniformly distributed in the network (over the long run), nodes that have the same radial distance from the processing center will deplete their energy supply approximately at the same time. From Section III, it is clear that the nodes within the first layer will run out of energy first. Furthermore, the portion of nodes that are in the first layer is small (approximately $\frac{1}{R_{net}^2}$). Clearly if there is no node in the first layer then the network will fail and cannot deliver information to the data sink anymore. Therefore, the lifetime of the first node to die in the first layer is closely related to the network lifetime.

With our analysis in Sections III and V, we are able to predict the peaks of energy consumption in the network. Every node has two components of energy consumption which contribute to the total energy consumption, e_t , of that node. The first part, e_{in} , is the average energy consumption due to being in a cluster either as a CH or cluster-member. This is the same for all nodes in the network, because each node is selected as a CH with the same probability (we neglect the edge effects, and clearly by CH updating, the load of being a CH is rotated periodically). The second part, $e_{out}(r)$, is the average energy consumption due to routing data toward the data sink, which is dependent on the distance, r , from the data sink as shown previously. Clearly, e_{in} , is the total amount of energy consumption within a cluster divided by the total number of nodes, n_{in} , in a typical cluster. As a first order approximation,

$$e_{in} \simeq \frac{\mathbf{E}S_{\lceil \frac{r}{R} \rceil}}{E(n_{in})}.$$

From [16] we have

$$E(n_{in}) \simeq \frac{\lambda_0}{\lambda_1} + 1 = \frac{1}{p}.$$

Combining the above with (9), we have

$$\begin{aligned}
e_{in} &\simeq p \frac{\lambda_0}{\lambda_1} \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2) \\
&= (1-p) \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2). \quad (13)
\end{aligned}$$

In the analysis of Section III, we have assumed that every node generates a new sensed data packet in each time unit with probability one. Clearly, if all nodes generate a packet with a probability p_0 , then their average energy consumption will be scaled by a factor of p_0 .

In our clustering model, the second component of energy consumption comes from forwarding the compressed data originated from the CHs. Since the point density of CHs is $p\lambda$, and each CH has $E(R_D(c_i))$ bits to send, the effective density of data is $p\lambda E(R_D(c_i))$. This means that the energy consumption curve will be scaled by a factor of $pE(R_D(c_i))$. Consequently, we have

$$e_{out}(r) = pE(R_D(c_i))E_c(r). \quad (14)$$

We combine (13) and (14) to obtain

$$\begin{aligned}
e_t(r) &= e_{in} + e_{out}(r) \\
&\simeq (1-p) \sum_{k=0}^{\lfloor \frac{R_{net}}{R} \rfloor} \exp(-\lambda_1 \pi (kR)^2) \\
&+ pE(R_D(c_i))E_c(r). \quad (15)
\end{aligned}$$

As can be seen from (15), given the characteristics of the correlation matrix we can compute the average $E(R_D(c_i))$ and then $e_t(r)$. As we have mentioned earlier, the critical nodes are at the verge of the first layer where $r = 1$. By definition, the network lifetime is proportional to the inverse of the maximum of energy consumption, i.e.,

$$\begin{aligned}
E(\text{Lifetime}) &\propto \min\left(\frac{1}{e_t}\right) \\
&= \frac{1}{e_t(1)}. \quad (16)
\end{aligned}$$

We define (16) as the *lifetime factor*. By computing (12) and (16) we can depict the total energy consumption versus network lifetime.

VII. NUMERICAL ANALYSIS AND SIMULATION

In this section, we provide simulation results as well as numerical analysis based on the the previous sections. In addition to comparing simulation results, we demonstrate the inherent trade-off between network lifetime and the total energy consumption. We also study the effect of data correlation and the amount of distortion on total energy consumption and network lifetime.

Our simulation environment consists of a square area of $20R \times 20R$ in which a total of $N = 2500$ nodes are uniformly distributed. We also set $D_0 = 0.01$ unless otherwise stated. We present results from both analysis and simulation for various values of parameter W , which is a representative of the amount of data correlation. Since very low values of p will result in too few clusters, we have limited p to be within an interval of $[0.002, 0.5]$.

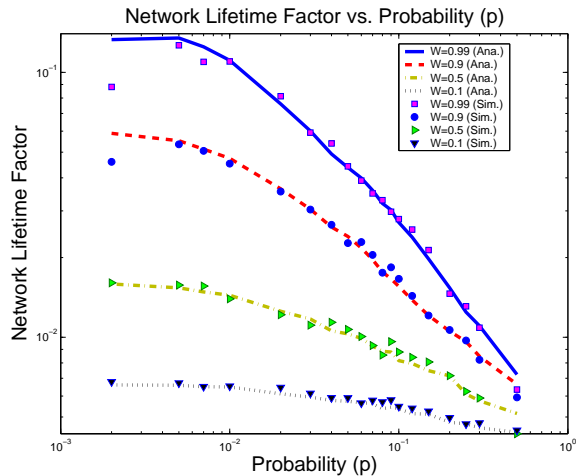


Fig. 4. Network lifetime vs. cluster-head probability ($D_0 = 0.01$)

A. Comparison with Simulation Results

Fig. 4 shows the network lifetime as a function of p . It can be seen that our analysis accurately predicts the behavior of network lifetime. Here our analysis is compared with the simulation results of clustering with MHR. The main reason of the good match between simulation and analysis in terms of network lifetime is that, from the point of view of nodes in the first few layers, the number of packets that should be relayed by them determines their energy depletion rate. Therefore, inaccuracies in the path lengths do not drastically influence the energy consumption analysis of the first or the second layer nodes.

As could be expected, increasing W increases the lifetime of network by more than an order of magnitude. This has been indicated but not analyzed in the past literature. Also, increasing cluster sizes can increase the network lifetime but as Fig. 4 suggests, after a point the curves are effectively flat and there is no more gain. The discrepancy between analysis and simulation when p is small is due to the finite size of the network and inadequate experimental data when there are few clusters.

In Fig. 5 an ideal routing scheme is considered, where a node that is at distance r from a destination needs $\lceil \frac{r}{R} \rceil$ hops to send its packets. Although this assumption may not be practical, it has been used as a guideline in the past literature especially when analyzing energy consumption [14]. As can be seen from the results, our analysis matches very well with simulation. In extreme cases where p is less than 0.005 or when it is near 0.5, the discrepancy is pronounced because of various approximations that we have made. For example, in either case, $\lceil \frac{r}{R} \rceil$ is not an accurate estimation. Also, when the clusters are large, the finite size of our network can affect the analysis.

Fig. 5 also compares our analysis with the simulation of clustering with MHR. As can be seen, the general behavior is the same except for approximately a 15% scaling factor. This is because $\lceil \frac{r}{R} \rceil$ is a lower bound to the number of hops to reach the destination and we expect longer paths for MHR as the density of nodes decreases. Despite the fact that there is

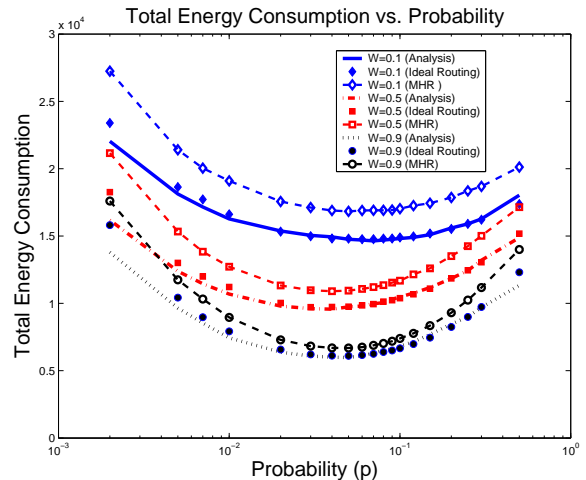


Fig. 5. Total energy consumption vs. cluster-head probability ($D_0 = 0.01$)

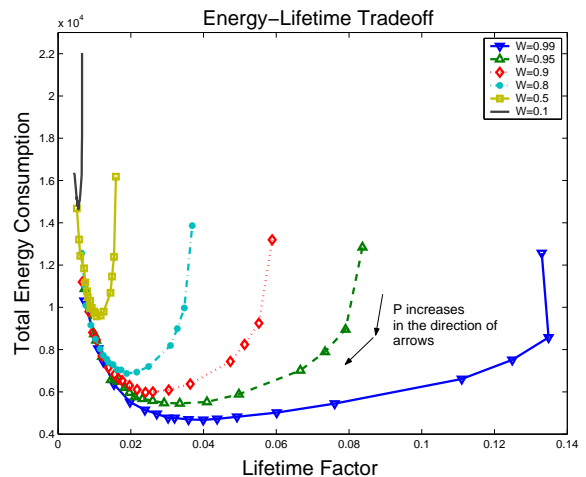


Fig. 6. Trade-off between energy consumption and lifetime. Each point corresponds to a value of p in [.002 .005 .007 .01 .02 .03 .04 .05 .06 .07 .08 .09 .1 .12 .15 .2 .25 .3 .5]. The direction by which p increases is indicated by arrows. Here $D_0 = 0.01$.

a scaling factor between two curves, the optimal cluster-head probabilities match very well.

B. Trade-off between Total Energy Consumption and Network Lifetime

Fig. 6 plots the total energy consumption against network lifetime for various values of p and W . Since the analysis and simulation match well, we show only the analysis results. This figure shows that we can trade-off the total energy consumption for the network lifetime. Clearly, in applications where sensors with non-renewable batteries are used, the latter is much more important than the former.

In all cases the behavior for all values of W is similar, but the optimal cluster-head probability and the performance gain are functions of W . For example, when $W = 0.99$, decreasing p from 0.05 to 0.005 will lead to more than 3 times

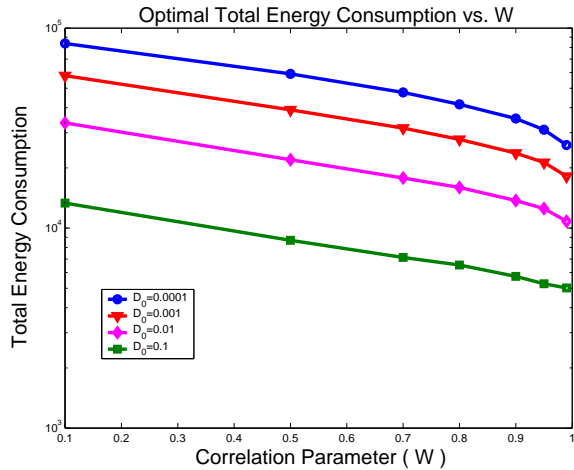


Fig. 7. Optimal total energy consumption vs. correlation parameter (W)

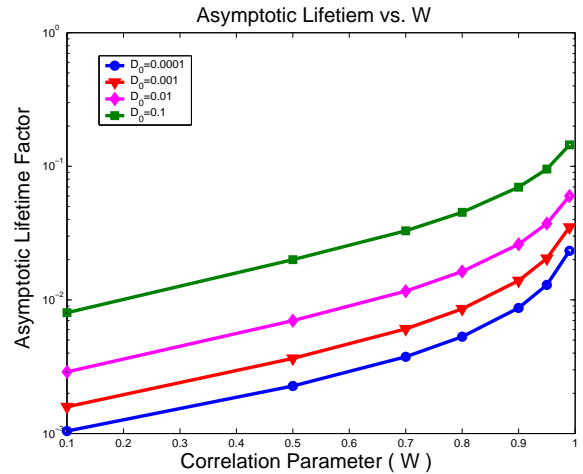


Fig. 8. Asymptotic lifetime factor vs. correlation parameter (W)

TABLE I

OPTIMAL PARAMETERS FOR THE CLUSTERING ALGORITHM

W	$D_0 = 0.0001$	$D_0 = 0.001$	$D_0 = 0.01$	$D_0 = 0.1$
0.99	0.04	0.04	0.06	0.06
0.95	0.03	0.03	0.05	0.06
0.9	0.02	0.03	0.05	0.06
0.8	0.02	0.03	0.04	0.05
0.7	0.02	0.04	0.04	0.05
0.5	0.03	0.04	0.04	0.05
0.1	0.06	0.06	0.07	0.08

improvement in the network lifetime while it increases the total energy consumption by a factor of less than 2. In contrast, for $W = 0.5$, the similar lifetime gain is less than 80% while the energy consumption is increased by a factor of about 70%.

C. The Effect of Data correlation and Compression Distortion

Fig. 7 and Fig. 8 show the effect of W and D_0 on the performance of clustering. Note that to compare the energy consumption and lifetime of the network for different values of D_0 , we need to scale (12) and (15) by $0.5 \log(\frac{1}{D_0})$.

Clearly, a low level of data correlation leads to inefficient clustering. These figures further suggest that the efficiency of clustering, as reflected by the minimal total energy consumption and asymptotic lifetime factor, increases with W at faster than exponential rate.

These figures also show that as the required amount of accuracy increases, i.e., D_0 decreases, the total energy consumption increases and the network lifetime decreases. This is an intuitive result, because increasing accuracy requires more bits used to represent the aggregated data. Furthermore, despite the different values of D_0 , the general behavior is similar. In fact, in log scale the different curves are biased approximately by a constant value. This suggests that energy and lifetime are directly affected by the choice of D_0 .

Table I shows the optimal cluster-head probability, which corresponds to the optimal cluster size, in terms of minimal total energy consumption for different values of W and D_0 . It can be observed that, within a certain range, as W decreases and D_0 is fixed, the optimal cluster-head probability first decreases, which means that to gain from data correlation we need more nodes in a cluster. When W is decreased further, to maintain a reasonable rate of compression, we need many samples or equivalently very large cluster sizes. However, large cluster sizes force the packets to travel long paths toward the CH, possibly away from the data sink, and at the same time because of inefficient compression, a large amount of packets need to be forwarded by the CHs. Therefore, many data will be forwarded through longer routes compared with the case where smaller clusters are used, which may negate the advantage of data compression. Consequently, the optimal cluster sizes decrease, and the optimal cluster-head probability increases.

Finally, for a fixed value of W , as D_0 decreases, the optimal cluster-head probability decreases. This suggests that the more stringent is the accuracy requirement, the larger is the optimal cluster size. Our numerical results indicate that the optimal cluster-head probability is approximately proportional to the logarithm of the distortion value.

VIII. RELATED WORK

Many clustering algorithms and protocols have been proposed in the past to improve the scalability of multi-hop wireless networks. They include single-hop clustering, first introduced in [6] and [7], and multi-hop clustering, for example [8] and [9]. Most of these algorithms do not consider energy consumption or network lifetime. With few exceptions, energy-aware clustering algorithms have been proposed mostly in the context of wireless sensor networking. They include [20], [21], [10], [11], [12], [13], [14], and [15]. Particularly, in [12][14], it has been noted that there exists an optimal number of clusters that minimizes total energy consumption. In [22] and [23],

the optimal number of databases, which correspond to cluster-heads, and their optimal arrangements have been derived for location and resource management. However, none of these algorithms consider the performance of data aggregation based on various data correlation levels.

Various distributed signal estimation protocols have been proposed for sensor networks. In [24], a distributed estimation algorithm is proposed for a subclass of periodic aggregation problems in which the result of aggregation is determined by the values of a few nodes. In [25], a distributed and adaptive signal processing algorithm is used to reduce the energy consumption. In [26], data funneling is proposed, in which border nodes of a queried region do the data aggregation and forward data to the sink. None of these works studies clustering or its effect on signal estimation.

In [27], the lifetime of a heterogeneous single-hop clustered network is analyzed, where the cluster-heads are high-capacity nodes that communicate directly with the data sink. In our work, we consider a flat network architecture, where all nodes have the same transmission power and communicate through multi-hop routing.

All clustering protocols discussed so far either do not consider data correlation or assume ideal aggregation, where an arbitrary number of packets within a cluster can be compressed into one packet. The ideal aggregation assumption is not valid in most applications. Except for the case of averaging or taking extrema, typically we need to observe data samples over an entire measured signal field. In this case, the objective is to provide a reasonable estimate of the signal field at any arbitrary point in the network.

In this work, we provide a novel framework to analyze the energy and lifetime performance of clustering algorithms, with a realistic routing protocol, under general data correlation functions and arbitrary compression distortion constraints. To the best of our knowledge no previous work has presented similar mathematical analysis on the effects of non-ideal aggregation on clustering in multi-hop sensor networks.

IX. CONCLUSIONS

In this paper, we have investigated the effect of partially correlated data on the performance of clustering algorithms. Using results from random geometry theory, we have studied the behavior of data forwarding with Minimum Hop Routing and demonstrated the fact that energy consumption is highly dependent on sensor location. Furthermore, such dependency is irregular, such that some nodes consume more energy even when they are farther away from the data sink than the others. To analyze network energy consumption, we have applied rate distortion theory to obtain lower bounds on the number of bits required to represent data within a cluster. We have further analyzed the network lifetime. We have shown that there exists a trade-off between total energy consumption and network lifetime, and that, in general, minimizing total energy consumption is far from optimality in terms of network lifetime. Our analysis framework provides a means to determine the optimal tuning

of the cluster-head probability to balance this trade-off based on application demands and hardware characteristics, as well as the cost of sensor battery replacement.

In addition, we have studied the dependency of energy consumption and network lifetime on the amount of data correlation and compression distortion. It has been concluded that data correlation affects the optimal energy consumption and asymptotic network lifetime faster than exponentially, and that compression distortion affects the optimal cluster-head probability logarithmically. In general, the lower is the data correlation, the less sensitive is the clustering algorithm performance, and the less is the requirement on compression accuracy, the smaller is the optimal cluster size to achieve minimum total energy consumption.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE communications magazine*, vol. 40, pp. 102–114, Aug 2002.
- [2] G.J.Pottie and W. Kaiser, "Wireless integrated network sensors," *Communications of The ACM*, vol. 43, no. 5, pp. 892–900, May 2000.
- [3] C. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of The IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug 2003.
- [4] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [5] A. Scaglione, "Routing and data compression in sensor networks: Stochastic models for sensor data that guarantee scalability," *Intl. Symposium on Information Theory, ISIT2003*, June 29– July 4 2003.
- [6] D. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *Transactions on communications*, vol. 29, no. 11, pp. 1694–1701, Nov. 1981.
- [7] A. Ephremides, J. E. Wieselthier, and D. J. Baker, "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proceedings of the IEEE*, pp. 56–73, Jan. 1987.
- [8] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max-min d-cluster formation in wireless ad hoc networks," *INFOCOM*, pp. 32–41, March. 2000.
- [9] B. Liang and Z. Haas, "Virtual backbone generation and maintenance in ad hoc network mobility management," *INFOCOM*, pp. 1293–1302, March 2000.
- [10] C. F. Chiasserini, I. Chlamtac, P. Monti, and A. Nucci, "Energy efficient design of wireless ad hoc networks," *In Proceedings of European Wireless*, Feb. 2002.
- [11] M. Chatterjee, S. k. Das, and D. Turgut, "Wca: A weighted clustering algorithm for mobile ad hoc networks," *Journal of Cluster Computing*, no. 5, pp. 193–204, 2002.
- [12] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor network," *IEEE Trans. on Wireless Communications*, vol. 1, pp. 660–670, Oct 2002.
- [13] A. Manjeshwar, Q. Zeng, and D. P. Agrawal, "An analytical model for information retrieval in wireless sensor networks using enhanced apton protocol," *IEEE Trans. on Parallel and Distributed Systems*, vol. 13, no. 12, pp. 1290–1302, Dec 2002.
- [14] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," *INFOCOM*, vol. 3, pp. 1713–23, March/April 2003.
- [15] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach," *IEEE INFOCOM*, March 2004.
- [16] S.G.Foss and S. Zuyev, "On a voronoi aggregative process related to a bivariate poisson process," *Advances in Applied Probability*, vol. 28, no. 4, pp. 965–981, 1996.
- [17] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression*, ser. Prentice-Hall series in information and system science. Englewood Cliffs, N.J. Prentice-Hall, 1971.

- [18] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," *IEEE/ACM Trans. Networking*, vol. 11, no. 1, pp. 2–16, Feb 2003.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley, 1991.
- [20] A. Cerpa and D. Estrin, "Ascent: Adaptive self-configuring sensor networks topologies," *INFOCOM*, pp. 1278–1287, June 2002.
- [21] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *ACM Wireless Networks*, vol. 8, no. 5, September 2002.
- [22] Z. J. Haas and B. Liang, "Ad hoc mobility management with uniform quorum systems," *IEEE/ACM Transactions on Networking*, vol. 7, no. 2, pp. 228–240, April 1999.
- [23] J. Li, Z. J. Haas, and B. Liang, "Performance analysis of random database group scheme for mobility management in ad hoc networks," *Proceedings of the IEEE International Conference on Communications (ICC)*, May 2003.
- [24] A. Boulis, S. Ganeriwal, and M. B. Srivastava, "Aggregation in sensor networks: An energy accuracy trade-off," *Proceedings of the First IEEE 2003 IEEE International Workshop*, pp. 128–138, May 2003.
- [25] J. Chou, D. Petrovic, and K. Ramachandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," *INFOCOM*, vol. 2, pp. 1054–1062, March/April 2003.
- [26] D. Petrovic, R. C. Shah, K. Ramchandran, and J. Rabaey, "Data funneling: routing with aggregation and compression for wireless sensor networks," *IEEE International Workshop on Sensor Network Protocols and Applications*, pp. 156–162, May 2003.
- [27] E. Duarte-Melo and M. Liu, "Analysis of energy consumption and lifetime of heterogeneous wireless sensor networks," *GLOBECOM '02*, vol. 1, pp. 21–25, 2002.